JOANNA KOMOROWSKA-MACH*

INTROSPECTION — ONE OR MORE?
PLURALISM ABOUT SELF-KNOWLEDGE**

Abstract

The aim of this paper is to provide a metaphilosophical analysis of the debate between monistic and pluralistic approaches to self-knowledge. I start by presenting the background for the discussion, clarifying the terminology, and drawing a distinction between moderate and radical pluralism. I then discuss and evaluate arguments for monism that appeal to general methodological guidelines as well as those relying on specific assumptions about the objects and character of self-knowledge. I offer a separate analysis of Eric Schwitzgebel's radical pluralism, which threatens to undermine both monism and moderate pluralism. I close by describing a way of looking at the problem of the homogeneity of self-knowledge that goes beyond the monism–pluralism dichotomy and arguing that the heterogeneity and multidimensionality of the problem of self-knowledge are more important than the plurality of the solutions.

Keywords: self-knowledge, introspection, pluralism, mental states, philosophy of mind

The question of whether our knowledge of our own mental states is always acquired in the same manner is a constantly recurring theme in contemporary debates over self-knowledge. Traditionally, it was believed that there is a universal source of self-knowledge about all kinds of mental states. This source, called introspection, was classically construed as a kind of inner perception (Locke 1690/1975, Russell 1997, Armstrong 1968, Goldman 2006, Gertler 2012). The proponents of alternative approaches toward self-knowledge, such as expressivism (Finkelstein 2003, Bar-On 2004), also aim to provide a model that would cover all the kinds of self-knowledge.

 * Institute of Philosophy, University of Warsaw, Krakowskie Przedmieście 3, 00-927 Warsaw, e-mail: j.komorowska-mach@uw.edu.pl, ORCID: 0000-0002-8287-6668.

However, there has been an increasing number of approaches in recent years that posit more than one first-personal way of obtaining self-knowledge. Such approaches are often called pluralistic (Schwitzgebel 2016). Moderate versions of pluralism postulate two or more sources of self-knowledge for several types of mental states (Coliva 2016), several modes of access to one's mental states (Renero 2018), or both (Boyle 2009). Radical versions of pluralism assert that there is a wide range of processes contributing to the acquisition of self-knowledge, that they need not be domain-specific (Prinz 2004), or even that particular instances of self-knowledge acquisition may consist in a "confluence of many processes, recruited opportunistically" (Schwitzgebel 2012: 29).

The aim of this paper is to provide a metaphilosophical analysis of the debate between monistic and pluralistic approaches toward self-knowledge. I start by presenting the background for the discussion, clarifying the terminology, and drawing a distinction between moderate and radical pluralism (part 1). I then discuss and evaluate arguments for monism that appeal to general methodological guidelines as well as those relying on specific assumptions about the objects and character of self-knowledge (part 2). In part 3, I offer a separate analysis of Eric Schwitzgebel's radical pluralism, which threatens to undermine both monism and moderate pluralism. I close by describing a way of looking at the problem of the homogeneity of self-knowledge that goes beyond the monism–pluralism dichotomy and arguing that the heterogeneity and multidimensionality of the problem of self-knowledge are more important than the plurality of the solutions (part 4).

1. WAYS OF GAINING SELF-KNOWLEDGE:
THE PROBLEM OF UNITY

All the positions discussed in this paper aim to explain the phenomena associated with self-knowledge. But what precisely those phenomena are and what features they possess is itself a matter for debate. For the time being, I am going to assume that a position falls within the scope of our discussion if it seeks to provide an answer to the following questions: (1) what is the source of one's judgments concerning one's occurrent mental states? and (2) why is it that utterances expressing mental self-ascriptions enjoy a special status that other kinds of utterances clearly lack (e.g., there seems to be nothing special about non-mental, third-personal utterances, or about utterances couched in the past tense)?

Accordingly, I take it for granted that, first, we issue judgments, aloud or in the head, about our present mental state ("I feel sad," or "I believe it is going to rain") and, second, that such utterances, which I call self-ascriptions, are generally taken at face value and rarely called into question or corrected. Following Donald Davidson (1984), I call this special status attributed to self-ascriptions *first-person authority*. I use this term to refer to an observable phenomenon, remaining neutral as to whether or not the special status accorded to self-ascriptions has any epistemic grounds. Traditional approaches explain first-person authority by postulating the existence of introspection — a hypothetical process or ability that allows one to "see" or detect one's present mental state in an epistemologically privileged way.[1] However, I will also take into account a variety of alternative approaches, such as expressivism, inferentialism, or constitutivism, that criticize the traditional introspectivist way of thinking about first-person access to mental states.

1.1. THE UNITY OF SELF-KNOWLEDGE

The monist thesis that there is a single process by which we gain first-person knowledge about all kinds of mental states is the default view. The position is not only popular but also grounded in philosophical tradition. Classical writings on self-knowledge, beginning with those by Augustine of Hippo, René Descartes, and John Locke, assume both unity and universality of self-knowledge. They assert that there is only one way of acquiring knowledge about one's occurrent mental states and that every occurrent mental state can be the object of self-knowledge.

These days, the universality assumption is rarely accepted. Psychologists, cognitive scientists, and contemporary analytic philosophers of mind use the term "mental state" in a broad sense that covers not only conscious experiences but also different kinds of states inaccessible to the subject. The term may refer to both dispositional features, such as personality traits, and subpersonal states implicated in information processing, such as Marr's 2,5D sketches (Smithies 2012). Therefore, a contemporary analogue of the universality assumption does not assert that every mental state can be self-ascribed but rather that every kind of mental state that can be self-ascribed can be self-ascribed in the way postulated by a particular approach.

---

[1] Many philosophers use the term "introspection" more broadly: i.e., to refer to all possible ways of gaining self-knowledge (e.g., Schwitzgebel 2012, Butler 2013). However, as I show elsewhere (Komorowska-Mach 2015a), this terminological decision may lead to equivocation.

By the homogeneity of self-knowledge, I mean the claim that there is only one way of gaining self-knowledge for all kinds of mental states knowable by the subject. One can take the universal approach without committing oneself to a homogeneity thesis; such a view might cover all self-ascribable mental states but posit two distinct sources of self-knowledge. One can also accept the homogeneity thesis without committing oneself to the universality assumption: e.g., by proposing one source of self-knowledge but focusing on a single type of mental state. Like universality, homogeneity is rarely considered to be absolute. It is widely accepted that we can also self-ascribe mental states in the same way we ascribe mental states to others — e.g., by analyzing behavior. Moreover, most views asserting that we obtain self-knowledge in a homogenous way allow alternative explanations in atypical cases.

The approaches that are both universal and homogenous (in the moderate senses of these two words described above) will be called monist approaches. A monist claims that, for all the main kinds of mental states recognised by folk psychology (including beliefs, thoughts, emotions, and feelings), it is possible to self-ascribe a state of this kind in the same common way generally associated with self-knowledge. Most introspective approaches, including Bertrand Russell's (1997), David Armstrong's (1968), William Lycan's (1996), Alvin Goldman's (2006), and Brie Gertler's (2012), satisfy this description. Moreover, some philosophers who question the existence of introspection, such as Dorit Bar-On (2004), Peter Carruthers (2011), and Quassim Cassam (2014), are also monists, as their approaches also apply to all major types of mental states.

Monism is a descriptive claim. Yet some of its critics, such as Matthew Boyle (2009) and Annalisa Coliva (2012, 2016), draw attention to its (often implicit) prescriptive aspect: accounts that are both universal and homogenous are regarded as more valuable than those combining more than one view. The "Uniformity Assumption," as Boyle puts it, asserts that "a satisfactory account of our self-knowledge should be fundamentally uniform, explaining all cases of 'first-person authority' in the same basic way" (Boyle 2009: 141). Coliva (2016) complains about a "monistic prejudice" that urges philosophers who are aware of their account's limited range of application to expand it artificially, at the expense of other assumptions, which finally undermines or weakens their position. This objection is raised against Bar-On's neo-expressivism, which according to some critics, provides a good explanation of how phenomenal states are self-ascribed but fails to account for self-ascriptions of intentional states (Komorowska-Mach 2015b, Coliva 2016).

1.2. TYPES OF PLURALISM

Pluralist approaches to self-knowledge postulate at least two different sources of self-knowledge, thereby rejecting the homogeneity thesis. In a moderate, "categorial" (Dąbrowski 2017) version, they also assume that we can distinguish certain stages or types of self-knowledge that are always realized by the same processes or mechanisms. For example, they might propose a different model of self-knowledge for each category of mental state, which often entails that self-ascriptions of distinct types of mental states differ in some respects.

The simplest version of this type of pluralism is the claim that there are two or more ways of acquiring self-knowledge and each can be assigned to some subset of mental states. Coliva's (2016) pluralistic account is a case in point, as it proposes three separate explanations for self-knowledge concerning three types of mental states: namely, phenomenal experiences, propositional attitudes, and dispositional states. By contrast, Boyle (2009) proposes to distinguish two types of self-knowledge: "active" and "passive," based on the properties of acts by which we can obtain it. The fact that this distinction partly overlaps with the dichotomy between phenomenal and intentional mental states is a secondary matter. Adriana Renero (2018) adopts a similar strategy, distinguishing stimuli-induced and self-triggered ways of acquiring self-knowledge. In this case, the division cuts across the taxonomy of mental states, although one might wonder whether each basic mental state is accessible in both these ways.

These three philosophers, all of whom explicitly use the term "pluralism" to define their approaches, postulate the existence of certain capabilities or processes dedicated to self-knowledge. However, pluralism can also be a starting point for an attempt at reducing self-knowledge to the abilities and processes responsible for other types of cognition. Jesse Prinz (2004) claims that, in fact, the term "introspection" denotes a set of different methods and processes which take part in accessing our internal representations. Our experience can be captured verbally, intensified as a result of focusing on it, reintegrated by combining information from different senses and using executive functions as well as recreated from memory. But none of these processes is specific to self-knowledge. Although Prinz does not rule out that "some unifying mechanisms might ultimately be found" (2004: 41), he shows the problematic character of the potential directions of such unification. He compares the term "introspection" to "natural disaster," arguing that similarities between different types of what we call self-knowledge may be shown only at a level of abstraction "too remote to support any substantive unification at the level of mechanism" (Prinz 2004: 48).

Pluralism in its radical, "structural" version (Dąbrowski 2017) asserts that every individual act of self-knowledge can be realized by a different set of processes and mechanisms (Schwitzgebel 2012). There are no substantial similarities between the different instances of self-knowledge acquisition, and even within one type, two particular instances can have entirely different foundations, components, and properties. Such a position is advocated by Schwitzgebel (who calls it "radical pluralism") and its goal is to show that there is no common ability or process dedicated to self-knowledge. When reconstructing arguments for and against pluralism, I will concentrate mainly on its variants which do not aim at showing that self-knowledge can be reduced to processes responsible for other abilities. One of the main reasons for this decision is the fact that among naturalistically oriented representatives of modern analytical philosophy of mind it is difficult to find someone who would question the thesis that there are many sub-personal processes and mechanisms involved in creating each belief, including those concerning one's own mental states. The fact that in ascribing to ourselves a mental state we use the resources of attention and memory, and the mechanisms and structures responsible for linguistic skills, etc. is uncontroversial even for the most conservative supporters of classical introspectivism.

Some authors go beyond the monism–pluralism dichotomy and intentionally restrict proposed explanations to a specific type of mental states (e.g., Davidson 1984, Moran 2001, Wright 1998). Sometimes their inclinations toward monism or pluralism are clear. For example, Richard Moran asserts explicitly that his account is intended to address the problem of how beliefs and intentions are self-ascribed and that a different model is needed to explain self-ascription of phenomenal mental states (Moran 2001: xxxiii). Other writers seem not to be concerned by the incompleteness of their models of self-knowledge. For example, Jordi Fernandez (2003), who has proposed an account of privileged self-knowledge of beliefs that is intended as an alternative to introspectivism, asserts that he focuses on belief for simplicity, though in the summary he seems to generalize his findings to all mental states as he claims that his approach allows us to eliminate "any mysterious faculty of introspection" (2003: 372). Furthermore, it is sometimes unclear whether an account is meant to apply to all types of mental states or not. For example, Gertler (2012) presents her concept of self-knowledge by acquaintance using only examples of such phenomenal states as pain and sensory experience.

2. ARGUMENTS CONCERNING THE UNITY OF SELF-KNOWLEDGE

Universality is taken to be an important constraint on accounts of self-knowledge (Bar-On 2004, Gertler 2010), but the reasons for this are never given. Indeed, it is standard practice to criticize a model of self-knowledge for not applying to a certain type of mental states, even if the model's author insists that the model is not intended to do so. This is how Bar-On (2004: 144) and David Finkelstein (2003: 155) criticize Moran's model, although he explicitly restricts the scope of its application to beliefs and intentions (Moran 2001: xxxiii).

Despite the intuitive appeal of monism, some philosophers abandon the search for a unified account of self-knowledge of mental states of all kinds. This is often because existing accounts fail to generalize. For example, Boyle (2009) claims that the explanation of first-person authority regarding beliefs provided by Moran (2001) is correct but incomplete. In effect, he argues that the ability to gain self-knowledge in a way postulated by Moran is fundamental as a necessary condition for possessing any self-knowledge at all. But at the same time, it clearly does not cover all types of acts in which we obtain self-knowledge. Therefore, according to Boyle, we should propose a second, "passive" way of getting self-knowledge over and above the "active" one described by Moran.

Other authors focus on the objects of self-knowledge. Coliva (2016) devotes much of her book to the differences between three broad types of mental states. Only then does she assign a different explanation to each type: constitutive for propositional attitudes, expressivist for feelings and emotions, and inferentialist for dispositional states.

Another pluralistic strategy is to show that some competing monistic approaches do not contradict but rather complement each other. Renero (2018) argues that two types of introspectivism: acquaintance and inner sense approaches, describe what are in fact two complementary methods of acquiring self-knowledge. Which of them is used depends on whether the act of gaining self-knowledge occurs involuntarily as a reaction to some stimuli or is brought about by an act of will and attention directing.

2.1. GENERAL METHODOLOGICAL ARGUMENTS

It may be possible to justify universal models of self-knowledge by appeal to some general methodological considerations. A general explanation for a large set of similar phenomena is preferred to a collection of different expla-

nations for particular subsets of the phenomena as both simpler and infor-
mationally richer. Furthermore, accounts that do not aim at universality may
be open to the charge of being ad hoc: offering different explanations for
cases that do not fit the proposed model.

However, the criterion of simplicity is problematic and unclear. Although
there are various motivations for preferring simpler hypotheses, laws, and
theories (for a short review see Baker 2016), justifying this preference is no
easy feat. For example, the simplicity of a thesis is often in conflict with its
informativeness (Lewis 1973: 73). Moreover, as there are different aspects of
simplicity, it is not possible to simply rank theories in terms of it. Structurally
simpler explanations can be more complex ontologically. And even ontologi-
cal economy itself can be understood in different ways: its qualitative and
quantitative criteria do not have to coincide (Lewis 1973: 87).

This problem is important for us because modern naturalistic explana-
tions of self-knowledge strive for a high ontological economy that sometimes
comes at the cost of structural simplicity. Classical introspectivism, on which
each act of gaining self-knowledge is explained by one type of ability, „sense,"
or mechanism, is structurally simpler than a complex pluralistic model like
the one proposed by Prinz (2004). On the other hand, Prinz's model is onto-
logically simpler because it does not postulate any additional entities apart
from those mechanisms and abilities that are already accepted in our expla-
nations of other human cognitive skills. Whether ontological simplicity is
more important than structural simplicity is disputable and dependent on
one's metaphysical and methodological assumptions.

Overall, general methodological considerations cannot provide a
straightforward solution to the problem under discussion. In the next three
sections, I reconstruct more detailed reasons for preferring monist models of
self-knowledge together with pluralist responses to them.

2.2. HOMOGENEITY OF THE OBJECTS OF SELF-KNOWLEDGE

What often underlies the preference for monist explanations of self-
knowledge is the assumption that mental states form a metaphysically ho-
mogenous class of objects. The idea that there is a realm of the mental whose
objects differ in important respects from other objects of cognition is not only
entrenched in folk psychology but also accepted and, to some extent opera-
tionalizable, in empirical sciences such as psychology and neuroscience.

In classical introspectivism, which serves as a model example of monism,
self-knowledge can be understood as a process similar to sense perception
but directed toward "internal" rather than "external" objects (Locke 1955).

This line of thought does not have to imply mind-body dualism. Some introspectivists unequivocally support materialism and ground the "internality" of the objects of self-knowledge by reducing them to the states of the nervous system (Armstrong 1968).

However, when applied to the objects of self-knowledge, the distinction between the internal and the external is dubious. First, speaking about mental states as inner states is a vague metaphor: there are lots of states which are located within the physical boundaries of the body or even its neural system, but have little to do with the domain of the mental.

Moreover, according to externalism, intentional mental states such as beliefs depend for their content on the external world (Putnam 1975, Davidson 1987). Therefore, externalistically construed beliefs cannot be internal states in the same sense as simple phenomenal states. This contrast may provide a strong reason for distinguishing two kinds of self-knowledge. We can draw a similar conclusion by appeal to active externalism (Chalmers, Clark 1998).

Another solution is to appeal to some functional properties of mental states and use them as a criterion of the internal. For example, we could use the property of conscious accessibility, which would imply that the distinguishing feature of mental states as is they come under the direct control of thought and action (Block 1997: 382). However, according to Smithies (2012), who has analyzed several functional criteria of the mental, it is impossible to give functional criteria that capture all and only those types of states that are usually taken to be potential objects of self-knowledge. He asserts that the only feature that singles out such objects is that they can be phenomenally individuated (Smithies 2012: 273). But that already assumes a universal concept of introspection (and also implies that beliefs have a particular phenomenology).

This is not the only line of argument for the heterogeneity of mental states. There are significant differences on many levels between types of mental states distinguished by folk psychology. It is hard to ascribe intentionality, the mark of the mental according to Franz Brentano (2009), to at least some phenomenal states (Rorty 1979, see also Wright 1998). Beliefs and intentions are analyzable in the language of behavioural dispositions, but such mental states as mental images are not. A subset of mental states, such as sensory experiences, seem to be associated with specific qualia, but the same can hardly be said about verbal thoughts or beliefs. We can ascribe pain or simple emotions to animals and discuss their adaptive functions without being accused of anthropomorphizing, but thoughts or intentions are considered exclusively human states. As Richard Rorty noticed:

> The attempt to hitch pains and beliefs together seems ad hoc — they don't seem to have anything in common except our refusal to call them "physical." (1979: 22)

Finally, most philosophers interested in the source of self-knowledge ignore some types of mental states because of their special properties. As Peter Langland (2015) shows, none of the available approaches explains how we obtain the knowledge of our imaginations. Mental images as well as other "borderline" cases of mental states, such as personality traits, are typically excluded from the discussion. It looks like we do not have "a unitary conception of what mentality consists in" (Kim 1996: 23) and the homogeneity of the objects of self-knowledge is no less controversial than the monist thesis itself.

Rorty (1970) has suggested that we do not assume our knowledge of mental states to come from a universal source because all mental states belong to the same category but rather we treat mental states as a homogenous class because we find that they are all accessible in a special way. If we accept that claim, all arguments for monism that invoke the homogeneity of the objects of self-knowledge are threatened by a vicious circle. We search for a universal model of self-knowledge because we believe that mental states form a homogeneous class, but at the same time we assume their homogeneity because we believe to have similar access to them.

However, supporters of the monistic approach toward self-knowledge can use this argument in their favor. It may be argued that homogeneity of self-knowledge is accepted not because we assume that mental states are homogeneous but because we notice that we perceive a particular group of states as homogeneous, which could be a sign that we use a similar "modality" towards them. Mental states — whatever they are — seem similar because their observable properties arise from the homogenous way in which our "internal sense" functions. Such reasoning can be illustrated with visual perception. We assume that our knowledge about colours of objects has a common source — perception — not because we assume any important similarity between the objects of perception (e.g., butterflies, Disney's movie, and a rainbow). Rather, given that we perceive most of objects as having colour, we assume that we have a similar access to them, regardless of whether colour is an objective property or colourful objects share important ontological similarities. By analogy, we can interpret the fact that all mental states seem to be similar in a particular way (internal, private, subjective, without spatial localisation, immaterial) as an argument in favour of the monist approach. Whatever such objects are and whichever objective properties they have, the fact that we perceive them as similar indicates that we have the same kind of access to them. Such argumentation, however, is not conclusive. Appeals to homogeneity do not seem to give a clear advantage to monism.

2.3. HOMOGENEITY OF THE ACTS OF GAINING SELF-KNOWLEDGE

One does not have to assume that mental states form a homogenous class of objects, or even believe that they are objects in any literal sense, to prefer monistic explanations of self-knowledge to pluralist ones. Properties of mental states aside, we can advocate for universal models of self-knowledge by appealing to the homogeneity of the acts by which we obtain it.

Homogeneity of the acts of gaining self-knowledge can be described at the phenomenological level. The fact that we use the word "introspection", both in psychology and in everyday language, to denote all such acts suggests that we intuitively perceive acts of gaining self-knowledge as relatively similar. Neither our language nor our experience reveals a tendency to feel or act differently when answering questions about different kinds of mental states, such as "What are you thinking about?," "How do you feel?," or "What do you want now?".

When discussing homogeneity of the acts of gaining self-knowledge, one may return to the internal–external metaphor described in the previous section. In both classical and contemporary literature (mostly by proponents of various kinds of introspectivism), we often encounter the statement that introspective acts themselves are "directed to the inside" of the subject (Armstrong 1968, Goldman 2006, Locke 1975). One can interpret this as a claim about the acts of gaining self-knowledge as well as a thesis about the objects of such knowledge.

Another feature that acts of self-knowledge seem to share, from the subject's perspective, is that from the point of view of a subject they seem to be direct at least in the sense that typically one does not need to use additional information or make inferences to self-ascribe mental state. That connects with the subjective feeling of certainty that accompanies self-knowledge. For example, Brentano claims that perception of mental phenomena is immediately evident (Brentano 1874/2009: 70).

As we can see, the acts of obtaining self-knowledge seem to be "directed" to the inside. However, this turns out to be misleading. Gareth Evans (1982) notes that, when we want to ascribe to ourselves mental states such as occurrent beliefs, we do not direct our attention "to the inside," at our own mental state, but rather "outside" — searching for information about the external world. To invoke a well-known example: when answering a question "Do you believe there will be World War III?" we are more interested in the objective state of the world than the state of our mind (Evans 1982, see also Burge 1988, Moran 2001). Additionally, Moran points out another difference between self-ascriptions of intentional and phenomenal states. He argues that we ap-

ply norms of rationality and responsibility to self-ascriptions of beliefs and intentions but not to utterances concerning emotions or feelings (Moran 2001, 2012).

Some philosophers claim that self-knowledge has a particular phenome-nology — that there is something *it is like* to be using introspection. For ex-ample, according to Lycan:

> When we attend to our own mental states it feels like . . . focusing our internal atten-tion. (2004: 101)

But this view is rather uncommon (see, e.g., Lyons 1986: 96, Shoemaker 1994, Siewert 2012). When obtaining self-knowledge about a phenomenal state we do not experience any sensations over and above those associated with being in that particular state. Some authors note that an experience may become intensified as a result of being the focus of attention (Prinz 2004), but the difference here is merely quantitative. Moreover, the introspecting of intentional states such as thoughts or beliefs does not seem to be accompa-nied by any qualia — the act of ascribing to oneself the belief that it is going to rain is not phenomenologically different from the act of ascribing to oneself the belief that it is going to be sunny. And Schwitzgebel (2012) discusses sev-eral cases confirming the claim that we do not have an impression of "doing the same thing" every time we self-ascribe a mental state.

Jakob Hohwy's (2013) model of introspection is an interesting variant of monism in that he explicitly raises the issue of an "introspective dissonance" in the phenomenology of self-knowledge; "introspection," he writes, "seems both accessible and certain, and inaccessible and uncertain" (2013: 247). The predictive account of the mind handles the heterogeneity inherent in the act of gaining self-knowledge by asserting that introspection is a species of at-tention phenomena and the dissonance in question results from precision optimization in prediction error minimization. As this example clearly shows, heterogeneity of the act of gaining self-knowledge need not undermine mo-nism, yet it is rarely properly addressed by its proponents.

2.4. HOMOGENEITY OF THE PRODUCTS OF THE ACTS OF GAINING SELF-KNOWLEDGE

It is also possible to defend the alleged superiority of universal explana-tions of self-knowledge by appealing to the homogeneity of what the acts of gaining self-knowledge produce: namely, beliefs or judgments. The special status of self-ascriptions can be analyzed at two levels (Davidson 1987). At the epistemological level, we can speak about the epistemic privilege of self-knowledge: e.g., infallibility or self-warrant. At the linguistic level, there is the

phenomenon of first-person authority: a tendency to treat utterances relating to one's own current mental states in a special way as compared with similar statements made from a third-person perspective. As Davidson remarks, philosophers who laid the ground for the modern understanding of introspection were primarily interested in the epistemic problem of self-knowledge (Davidson 1984). Alston discusses the main variants of epistemic privilege and, although critical of their strongest interpretations, he believes that a moderate version of epistemic privilege of self-knowledge is beyond doubt and that "no argument designed to show that this is false or meaningless can be sound" (Alston 1971: 224).

Even philosophers who are more skeptical of epistemic privilege of self-knowledge usually accept that utterances in which we self-ascribe mental states have special authority. The objects of self-knowledge may vary, yet first-person authority, as an observable social and linguistic practice, seems to apply to all the main types of mental states in much the same way. This is why Bar-On insists that a satisfactory concept of self-knowledge should explain the source and character of first-person authority, which — in her opinion — holds for self-ascriptions of all types of mental states, both intentional (such as beliefs and desires) and phenomenal (such as feelings and emotions) (2004: 20). Although she stresses that she does not want to deny the existence of "interesting differences" between phenomenal and intentional states, she argues that, in the end, both these classes enjoy a special status of the same type (Bar-On 2004: 5-6).

However, product-homogeneity of self-knowledge may be questioned on both epistemic and pragmatic grounds. Starting with the former, it is not clear whether we can obtain knowledge in the classical sense regarding all types of mental states. The fact that we can give accurate linguistic expression to current thoughts or beliefs seems uncontroversial, since we usually assume that they already come in propositional form. But when it comes to describing our feelings, sensations, or mental images, things get more complicated. One can focus on the experience of seeing an ice cube, memorize it, retrieve it from memory, and use as input for further mental operations, such as those involved in recognizing an object with one's eyes closed. But this is a knowledge of a different kind than that associated with, say, being certain that the ice cube in my hand will eventually melt (see Prinz 2004).

Reliability also seems to vary with the type of self-knowledge involved. William Seager (2002) claims that self-knowledge about emotional states is much less reliable than self-knowledge about other kinds of phenomenal states. In their well-known critique of classical introspectivism, Richard E. Nisbett and Timothy D. Wilson (1977) discuss a wide range of psychological

studies suggesting that people are even less reliable when they ascribe beliefs and attitudes to themselves.

At the pragmatic level, the focus is on the way utterances about our own mental states function in sociolinguistic practices. Even philosophers who claim that first-person authority is shared by all types of self-ascriptions agree that its strength varies according to the type of mental state. Crispin Wright (1998: 17) contrasts utterances in which we self-ascribe phenomenal states, such as sensations and emotions, with those in which we self-ascribe intentional states (beliefs, intentions). Whereas the former enjoy high authority, the latter are occasionally questioned and corrected. Moran additionally points out that why-questions addressed to a person who has ascribed to herself a mental state, are interpreted differently depending on whether the self-ascribed state is phenomenal or intentional. A person who has self-ascribed pain, will interpret the question as being about the causes of her state and her reply will likely invoke injury or illness. But the same question addressed to a person who has self-ascribed the belief that it is going rain would be interpreted as concerning the reasons for endorsing the belief rather than the causes of its adoption. Moreover, in the first case, failure to give an answer would be fully acceptable (we are allowed not know why we are in pain) and an answer provided by a third party (e.g., a doctor) may be considered more valuable than that given by the person concerned. In the second case, we are obligated to provide a reply and we should be able to explain why our belief is worth adopting; at the same time, our explanation has more weight than an answer supplied by a third party: "we take the person to be uniquely relevant as the person to ask regarding what is a reasonable or worthwhile in some attitude she holds" (Moran 2012: 215, see also Smithies 2012).

Depending on the perspective taken, one can emphasize either the similarities or the differences between the various kinds of self-knowledge produced. But again, none of those perspectives prevails when it comes to deciding whether we need distinct explanations for different kinds of self-knowledge. The question stated at the beginning remains open.

3. SCHWITZGEBEL'S RADICAL PLURALISM

In his provocative paper "Introspection, What?," Schwitzgebel (2012) presents his radically pluralist approach. He claims that most self-ascriptions arise from a multitude of processes ("within-case pluralism") and that the collection of processes that contribute to the formulation of each self-

ascription also differs from case to case ("between-case pluralism"). Not only does he deny monism but he also rejects moderate pluralism, which asserts that it is possible to capture self-knowledge using several complementing models. According to Schwitzgebel, every particular act of self-ascription, even of the same type and performed by the same person in a similar context, may have entirely different grounds and stages and exploit different processes, mechanisms, and abilities, recruited opportunistically (2012: 29).

The thesis itself is similar to that pursued by Jesse Butler (2013), who asserts that there are many processes involved in getting self-knowledge. But Schwitzgebel's conclusions regarding the problem considered in this paper are far more radical. According to Schwitzgebel, even Prinz's deeply reductivist version of pluralism turns out not to be sufficiently radical because it still assumes that there are types or stages of gaining self-knowledge which are systematically related to particular mechanisms or processes. Also, in contrast to Prinz, Schwitzgebel questions the possibility of empirical confirmation of any general model of self-knowledge. Alluding to authors such as Prinz or Shaun Nichols and Stephen Stich, who illustrate their account of self-knowledge using schematic diagrams (Nichols, Stich 2003), he proposes his own „boxology" — which is a tangle of lines — and claims that such "crazy spaghetti" is the most accurate model of the phenomena (Schwitzgebel 2012: 40).

As evidence for this radical thesis, Schwitzgebel presents a detailed description of three acts in which different types of mental states (a visual experience, an emotion, and an auditory imagination) are self-ascribed (all of them are presented as his own experiences). He reconstructs and enumerates the processes, abilities, actions, and mechanisms that contribute to these acts, focusing on both diversity and contingency of such components. These three case-studies are supposed to show that there is no process common to different acts of gaining knowledge — or even to their subsets. But the value of such evidence is questionable. Schwitzgebel's reconstruction of the stages and components of the acts of gaining self-knowledge are highly subjective, based on the first-personal perspective and formulated in the language of folk psychology. In other words, he investigates introspection by means of introspection. The problem is that even the most vehement supporters of classical introspectivism do not claim that we have infallible knowledge of subpersonal components of our conscious experience (see Nisbett, Wilson 1977). Schwitzgebel's reconstructions may be incorrect or incomplete. Therefore, his general conclusion that there are no processes or mechanisms common to all the acts of gaining self-knowledge is unjustified.

Moreover, Schwitzgebel completely ignores the phenomenon of first-person authority. His radical approach does not explain why self-ascriptions have

such a unique status in our socio-linguistic practice. The claim that introspection is just one among many "broad types of cognitive processes recognised by the terms of folk psychology" (2012: 40) is not an explanation — the very fact that folk psychology distinguishes such an ability demands an explanation.

There is one more problem with Schwitzgebel's approach. All the examples discussed in his paper rely on a very broad meaning of the word "introspection," roughly synonymous with "a way of gaining any kind of self-knowledge." This meaning, he claims, is accepted by "most philosophers." Yet on this interpretation of the term, Schwitzgebel's thesis becomes trivial. The fact that the term "introspection" is sometimes quasi-colloquially applied to phenomena that do not have a common cognitive basis does not imply that it is impossible to develop a scientifically useful theoretical model of self-knowledge based on a narrower technical concept of introspection. Likewise, the fact that the word "memory" is applied quite loosely in everyday language does not prevent cognitive psychologists from using the term, after it has been suitably defined, in scientific theorizing.

In the final section of his paper, Schwitzgebel proposes his own characterization of introspection:

> Introspection is the dedication of central cognitive resources, or attention, to the task of arriving at a judgment about one's current, or very recently past, conscious experience, using or attempting to use some capacities that are unique to the first-person case . . . with the aim or intention that one's judgment reflect some relatively direct sensitivity to the target state. (2012: 42-43)

It is unclear how to reconcile this characterization with the previous conclusion that no general definition or model of introspection is possible. Schwitzgebel suddenly admits that different acts of gaining self-knowledge have something in common and can be captured using theoretical concepts. By why not define "introspection" as these "capacities that are unique to the first-person case"? Even if Schwitzgebel's examples were to force us to reject monism, his next step, leading from categorial to radical pluralism, would not be sufficiently justified.

An accurate account of all the different types of self-knowledge would have to invoke many types of processes and abilities. However, the conclusions Schwitzgebel draws from his observations are too far-reaching. Radical pluralism does not provide a solution to the problems related to self-knowledge nor does speak against the rationality of searching for one.

4. PLURALISM OR A PLURALITY OF PROBLEMS?
IN SEARCH OF A DIFFERENT PERSPECTIVE

As I have shown, the reasons for choosing homogenous approaches to self-knowledge can be seriously undermined by pluralists. But the pluralists' argumentation is not conclusive and a radical version of pluralism, an attempt to dismiss (or even ridicule) the whole discussion, does not withstand criticism either.

However, a rejection of the prescriptive versions of both monism and pluralism is itself an important result. There are no grounds for insisting that only monism can provide a satisfactory explanation of the phenomena associated with self-knowledge, but then again, there are no grounds for insisting that monism should be rejected.

My reconstruction of the arguments for and against the homogeneity of self-knowledge casts the problem of self-knowledge in a new light. It makes it clear that the question about the source of self-knowledge is asked simultaneously at multiple levels. There is the psychological level: How is achieving self-knowledge possible? How does it work? What processes and mechanisms are responsible for it?, and there is the epistemological level: How are our beliefs about our own current mental states justified? Is self-knowledge privileged? But we also need to account for certain social practices: the asymmetries between self-ascriptions and third-person utterances. Finally, there is the moral aspect associated with the obligations and responsibility of the subject.

The approaches to self-knowledge discussed in this paper, both monist and pluralist ones, address only some, or just one, of these problems, which means that pitting them against one another is rarely genuinely informative. Before comparing available solutions in terms of how many sources of self-knowledge they posit, we should first look at their explananda — the phenomena that make up the "problem" of self-knowledge. This approach is pursued by Robert Piłat, who analyzes different dimensions of the problem itself, including the semantic, epistemic, normative, and ethical-existential level (Piłat 2013)

Unification would be possible if a single approach were able to respond satisfactorily to the problems investigated at all the levels. Such an approach does not seem to be available. Instead, individual philosophers argue that certain layers of the problem of self-knowledge are significant while others may be neglected, or that certain phenomena that seem to require explanation do not really exist.

Classical introspectivism is primarily a solution to the epistemological problem. It aims at explaining the special epistemic status of self-knowledge which itself accounts for first-person authority. If introspection is epistemically privileged and immune to error it is not surprising that self-ascriptions have enjoy a special status. But introspectivism does not answer the psychological question of how any cognitive process or mechanism can be direct and infallible. For those philosophers of mind who keep in touch with empirical findings, the classical solution is not convincing.

Materialistic introspectivism starts from an empiricist perspective, which is closer to the psychological approach. Armstrong postulates the existence of a special kind of "internal scanner" explaining the acquisition of self-knowledge. At the same time, he undermines the importance of the epistemic problem of grounding of self-knowledge and completely omits the pragmatic aspect: Armstrong's self-knowledge is as unreliable as perception so there is no reason for it to enjoy such special authority.

Bar-On's neo-expressivism focuses on the pragmatic level: it assumes the homogeneity and uniqueness of first-person authority as a special social practice. Bar-On tries to provide an explanation at the psychological level, emphasizing continuity between verbal and natural expression, but her explanation of the epistemological problem is criticized as unconvincing (Boyle 2010, Brueckner 2011, Rosenthal 2010).

Moran, on the other hand, is not interested in the psychological aspect of the problem of self-knowledge and whether his position offers an explanation of the epistemological problem is also questioned (Gertler 2010). However, he does focus on the normative aspects of self-knowledge overlooked by most philosophers and links the notion of self-knowledge with the notions of the rationality and responsibility of the subject.

Finally, Schwitzgebel's pluralism concentrates on the psychological question of which processes and mechanisms underlie self-knowledge and disregards the epistemological and pragmatic aspects of self-knowledge.

Recognizing the many-faceted nature of self-knowledge can lead to a re-evaluation of all the positions in the debate. Categorial pluralism may be supported by showing that particular aspects of the problem have different weights for each type of self-knowledge. And so, for example, epistemological aspects are particularly important for the problem of self-knowledge about occurrent thoughts, the psychological aspect is most pressing for the problem of phenomenal states, the pragmatic level is crucial when we talk about expressing emotions, and the normative level becomes important for the problem of beliefs and attitudes. For some types of self-knowledge, some of these levels may be less relevant than others.

Schwitzgebel criticizes monism using a well-known metaphor of blind men and an elephant. Those who hold the trunk think that the elephant is flexible and long, those closer to its fangs believe that it is hard and sharp at the end. But the problem of self-knowledge is even more complicated because philosophers who study it differ in their goals, expectations, research methods, and criteria of success. In addition, they are not dealing with an elephant, an observable, intersubjectively accessible physical object, but with a whole spectrum of phenomena, most of which are hypothetical objects.

This does not mean that, as Schwitzgebel suggests, we should give up the ambition of explaining phenomena related to self-knowledge in any systematic way. Rather, it means that even presenting the problem and understanding the hidden assumptions and the implicit aims of the solutions requires a multidimensional approach, which involves much more work than meets the eye. There is no reason to believe that the untangling of what Schwitzgebel mockingly calls a "cognitive confluence of crazy spaghetti" (2012: 41) is fundamentally impossible. We only must start from the right end of it.

BIBLIOGRAPHY

Alston W. (1971), "Varieties of Privileged Access," *American Philosophical Quarterly* 8(3), 223-241.

Armstrong D. (1968), *A Materialist Theory of the Mind*, London: Routledge.

Baker A. (2016), "Simplicity" [in:] *Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), E. N. Zalta (ed.), https://goo.gl/nW4avd.

Bar-On D. (2004), *Speaking My Mind: Expression and Self-Knowledge*, Oxford: Oxford University Press.

Block N. (1997), "On a Confusion about a Function of Consciousness" [in:] *The Nature of Consciousness*, N. Block, O. Flanagan, G. Guzeldere (eds.), Cambridge, MA: MIT Press, 375-416.

Boyle M. (2009), "Two Kinds of Self-Knowledge," *Philosophy and Phenomenological Research* 78(1), 133-164.

Boyle M. (2010), "Bar-On on Self-Knowledge and Expression," *Acta Analytica* 25(1), 9-20.

Brentano F. (2009), *Psychology from an Empirical Standpoint*, London—New York: Routledge.

Brueckner, A. (2011), "Neo-expressivism" [in:] *Self-Knowledge*, A. Hatzimoysis (ed.), Oxford: Oxford University Press.

Burge T. (1988), "Individualism and Self-Knowledge," *The Journal of Philosophy* 85(11), 649-663.

Butler J. (2013), *Rethinking Introspection: A Pluralist Approach to the First-Person Perspective*, Dordrecht: Springer.

Carruthers P. (2011), *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press.

Cassam Q. (2014), *Self-Knowledge for Humans*, Oxford: Oxford University Press.

Chalmers D., Clark A. (1998), "The Extended Mind," *Analysis* 58(1), 7-19.

Coliva A. (2012), "One Variety of Self-Knowledge: Constitutivism as Constructivism" [in:] *The Self and Self-Knowledge*, A. Coliva (ed.), Oxford: Oxford University Press, 212-242.

Coliva A. (2016), *The Varieties of Self-Knowledge*, London: Palgrave Macmillan.

Dąbrowski A. (2017), "Czym jest introspekcja? Badanie rozwiązań pluralistycznych," *Analiza i Egzystencja* 38, 39-66.

Davidson D. (1984), "First Person Authority," *Dialectica* 38(2-3), 101-111.

Davidson D. (1987), "Knowing One's Own Mind," *Proceedings and Addresses of the American Philosophical Association* 60(3), 441-458.

Evans G. (1982), *The Varieties of Reference. Philosophy* (vol. 2), Oxford: Clarendon Press.

Fernandez J. (2003), "Privileged Access Naturalized," *The Philosophical Quarterly* 53(212), 352-372.

Finkelstein D. H. (2003), *Expression and the Inner*, Cambridge, MA: Harvard University Press.

Gertler B. (2010), *Self-Knowledge*, London: Routledge.

Gertler B. (2012), "Renewed Acquaintance" [in:] *Introspection and Consciousness*, D. Smithies, D. Stoljar (eds.), Oxford: Oxford University Press, 93-127.

Goldman A. (2006), *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading. Neuroscience*, Oxford: Oxford University Press.

Hohwy J. (2013), *The Predictive Mind*, New York: Oxford University Press.

Kim J. (1996), *Philosophy of Mind. The Philosophical Review*, Boulder, CO: Westview Press.

Komorowska-Mach J. (2015a), "Pojęcie introspekcji w psychologii potocznej," *Przegląd Filozoficzny — Nowa Seria* NUM, 75-89.

Komorowska-Mach J. (2015b), "Przypisywanie sobie emocji — wyrażanie czy wykrywanie? Elementy detektywistyczne w modelu neoekspresywistycznym," *Filozofia Nauki* (*The Philosophy of Science)* 23(4) [92], 41-54.

Langland-Hassan P. (2015), "Self-Knowledge and Imagination," *Philosophical Explorations* 18(2), 226-245.

Lewis D. (1973), *Counterfactuals*, Oxford: Basil Blackwell.

Locke J. (1975), *An Essay Concerning Human Understanding*, Oxford: Oxford University Press.

Lycan W. G. (1996), *Consciousness and Experience*, Cambridge: MIT Press.

Lycan W. G. (2004), "The Superiority of HOP to HOT" [in:] *Higher Order Theories of Consciousness*, R. Gennaro (ed.), Philadelphia: John Benjamins Publishers, 93-113.

Lyons W. (1986), *The Disappearance of Introspection*, Cambridge: MIT Press.

Moran R. (2001), *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton: Princeton University Press.

Moran R. (2012), "Self-Knowledge, 'Transparency,' and the Forms of Activity" [in:] *Introspection and Consciousness*, D. Smithies, D. Stoljar (eds.), Oxford: Oxford University Press, 211-236.

Nichols S., Stich S. P. (2003), *Mindreading*, Oxford: Oxford University Press.

Nisbett R., Wilson T. (1977), "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84(3), 231-259.

Piłat R. (2013), *Aporie samowiedzy*, Warszawa: IFiS PAN.

Prinz J. (2004), "The Fractionation of Introspection," *Journal of Consciousness Studies* 7-8, 40-57.

Putnam H. (1975), "The Meaning of 'Meaning'," *Minnesota Studies in the Philosophy of Science* 7, 131-193.

Renero A. (2018), "Modes of Introspective Access: A Pluralist Approach," *Philosophia*, 1-22.

Rorty R. (1970), "Incorrigibility as the Mark of the Mental," *The Journal of Philosophy* 67(12), 399-424.

Rorty R. (1979), *Philosophy and the Mirror of Nature*, Princeton: Princeton University Press.

Rosenthal D. M. (2010), "Expressing One's Mind," *Acta Analytica* 25, 21-34.

Russell B. (1997), *Problems of Philosophy*, Oxford: Oxford University Press.

Schwitzgebel E. (2012), "Introspection, What?" [in:] *Introspection and Consciousness*, D. Stoljar, D. Smithies (eds.), New York: Oxford University Press, 29-47.

Schwitzgebel E. (2016), "Introspection" [in:] *Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), E. N. Zalta (ed.), https://goo.gl/47XmSb.

Seager W. (2002), "Emotional Introspection," *Consciousness and Cognition* 11(4), 666-687. Retrieved from https://goo.gl/Vkx95g.

Shoemaker S. (1994), "Self-Knowledge and 'Inner Sense.' Lecture I: The Object Perception Model," *Philosophy and Phenomenological Research* 54(2), 249-269.

Siewert C. (2012), "On the Phenomenology of Introspection" [in:] *Introspection and Consciousness*, D. Smithies, D. Stoljar (eds.), Oxford: Oxford University Press, 129-167.

Smithies D. (2012), "A Simple Theory of Introspection" [in:] *Introspection and Consciousness*, D. Smithies, D. Stoljar (eds.), Oxford: Oxford University Press, 259-327.

Wright C. (1998), "Self-Knowledge: The Wittgensteinian Legacy" [in:] *Knowing Our Own Minds*, C. Wright, B. Smith, C. Macdonald (eds.), Oxford: Oxford University Press, 101-122.